JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

DEVELOPMENT OF AN ARTIFICIAL INTELLIGENCE-SUPPORTED CHATBOT AS AN INTERACTIVE LEARNING PLATFORM IN STEM EDUCATION: EXPLORING USABILITY AND STUDENT EXPERIENCE

Gábor Fekets

Abstract. *While generative artificial intelligence (AI) tools are gradually being integrated into educational practice, their actual usability in classroom settings remains insufficiently understood. This mixed-methods research was designed as a small-sample pilot study, offering preliminary insights to inform a future large-scale scientific evaluation. The following questions were explored: What are the contents, functions, and capabilities of AI-based chatbots in education? How do students evaluate the usability of a chatbot using BUS-15 metrics? The online developer system botpress.com was chosen as the framework to support physics education within STEM fields for upper-secondary students. A validated instrument, BUS-15, a commercial tool for measuring chatbot usability, was utilised in an educational setting to evaluate students' user experience. In May 2025, 37 Taiwanese students used the developed AI chatbot. Cronbach's α was 0.709, consistent with values typically observed for chatbots developed for business purposes. Based on the results, the chatbot met the students' expectations. Accessibility emerged as a key determinant of chatbot success, although issues were noted with response times and conversational coherence. The findings also underscore the importance of evaluating a chatbot's pedagogical value, specifically its potential to enhance learning experiences and support multilingual engagement.*

Keywords: *AI chatbot, BOT Usability Scale, Botpress in education, BUS-15 questionnaire, interactive learning platform*

Gábor Fekets
National Taiwan Normal University, Taiwan

Introduction

The global COVID-19 pandemic accelerated demand in the educational sector for technologies that facilitate online learning. Artificial intelligence (AI) chatbots have emerged as one of the most widely adopted tools in this context (Okonkwo & Ade-Ibijola, 2021). These AI-driven applications support student learning by providing personalised instruction and real-time feedback, reducing the administrative burden (Bettayeb et al., 2024; Mishra & Varshney, 2024). Additional advantages include around-the-clock accessibility and human-like conversational interaction, both of which contribute to a more seamless educational experience (Ghayoomi, 2023). Research has also shown that AI chatbots enhance student motivation, engagement, and satisfaction in STEM education by improving service quality and delivering tailored support, factors that ultimately contribute to improved student retention (Neji et al., 2023; Segovia-García, 2024).

Previous studies have highlighted the pivotal role of ChatGPT in AI-supported learning pathways, with particular success in personalised instruction and real-time assessments in mathematics education (Opesemowo & Adewuyi, 2024). Performance expectations strongly influence users' intended outcomes, which in turn affect continued use of ChatGPT (Bazelais et al., 2024). Saxena and Doleck (2023) reported rapid acceptance of the platform, reinforcing the need to understand the factors driving user intention and sustained engagement. Wardat et al. (2023) examined the integration of ChatGPT into mathematics teaching, confirming its positive impact on student success through comprehensive guidance and instructional support. In broader STEM education, chatbots have demonstrated the ability to enhance student confidence and learning outcomes by delivering timely, adaptive assistance (Luzano, 2024). Their effectiveness is particularly evident in science instruction, where they outperform traditional rule-based systems in fostering active learner participation (Ng et al., 2024).

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

A recent scoping review examined students' perceptions of AI chatbots in higher education, emphasising their motivational benefits and overall usefulness, while also raising concerns related to response accuracy and potential negative effects on learning (Schei et al., 2024). Notably, researchers have cautioned against diminished creativity and critical thinking skills, suggesting the need for further analysis of the broader cognitive impacts of chatbots (Hasan et al., 2024). Some concerns suggest that increased reliance on chatbot-generated responses may erode the role of human instructors and the teacher–student dynamic, with some studies noting potential emotional attachment to AI systems. A key research gap lies in the limited availability of locally relevant knowledge bases, which can undermine the accuracy and contextual relevance of chatbot responses. Addressing this limitation is essential for improving both the effectiveness and adaptability of AI chatbots in educational settings (El Azhari et al., 2023).

At the same time, the integration of AI-supported tools in education brings ethical and legal challenges, particularly regarding data privacy and the preservation of meaningful human interaction in the learning process (Hasan et al., 2024; Schei et al., 2024). Rule-based chatbots have become less effective due to the rapid advancement of computer performance and natural language processing (NLP). The revolution in AI and generative models has significantly enhanced the conversational capabilities of chatbots and their ability to provide complex responses.

Research Problem

This research addresses the following three challenges. First, student disengagement in physics, which is often viewed as complex and uninspiring, can negatively impact academic performance and diminish enthusiasm for future engagement in STEM fields. Second, the study introduces an AI-based chatbot specifically designed to foster engagement and evaluate user experience through statistical methods. By leveraging interactive features and personalised support, the chatbot aims to create a more stimulating and accessible learning environment. This innovative approach is intended to boost student interest in physics, deepen conceptual understanding, and cultivate a more positive attitude toward STEM education. Third, despite the rich theoretical landscape surrounding chatbot development, no commercial instruments have yet been employed to assess user experience within educational settings. The PARADISE framework (Walker et al., 1997) supports performance measurement in spoken dialogue systems, offering a comprehensive tool for evaluating conversational accessibility, functionality, interaction quality, reliability, and user trust. While usability tools such as Usability Metric for User Experience (Finstad, 2010) are commonly used to gauge system and interface usability, they do not explicitly address chatbot-specific characteristics. Valério et al. (2018) also emphasised the absence of instruments that capture end-user experience in chatbot interactions. As chatbot applications grow within commercial domains, the need to evaluate their qualitative performance has become increasingly urgent (Federici et al., 2020). Responding to this gap, Borsci et al. (2022) developed the BOT Usability Scale (BUS-15), a 15-item questionnaire with demonstrated reliability ranging from .76 to .87, designed to assess user satisfaction with AI conversational agents.

To the best of my knowledge, no previous studies have implemented a customised AI chatbot while measuring user experience in the educational domain. Several research gaps remain: for example, no studies have assessed tools or development methods for implementing customised educational chatbots; chatbots are typically deployed as standalone tools rather than integrated systems, and no studies have used AI to assess student performance. Additionally, no generalised model currently exists to define essential chatbot functions and services, and standardised usability instruments and statistical methods for evaluating educational chatbot interactions remain absent.

To address these gaps and advance research in this domain, a customised AI-based chatbot was developed using the botpress.com developer platform, and its user experience was evaluated using standardised quantitative instruments.

Research Focus

As generative AI tools have become increasingly affordable and accessible, questions have emerged regarding their practical application and effectiveness within the education sector. This research addresses the following key areas. First, a review of the state of the art in AI chatbot technologies is conducted. Second, the application of chatbots in educational contexts is explored. Third, an integrated chatbot platform is described, detailing its functions, data structures, and defined user roles and tasks for students, teachers, and parents. Finally, the platform is evaluated in terms of its usability in upper-secondary education.

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Research Aim and Research Questions

First, a new AI-based chatbot tool using the botpress.com development framework is presented. Then, its usability is assessed based on students' user experiences, with a focus on exploring how AI-driven chatbots can support students in STEM education. Two research questions are posed: What content and functions should AI-driven chatbots offer to support students in STEM education? Do students find such chatbots useful?

Research Methodology

General Background

The continued development of AI has opened new possibilities for personalised learning. Chatbots serve as interactive tools that enhance student engagement and support efficient knowledge retrieval. This study developed a customised AI chatbot and tested its usability and student satisfaction in a physics course for upper-secondary students. Numerous recent studies have explored the use and impact of chatbots in education. However, the user experience remains underexplored, with most research focusing solely on student satisfaction, rather than usability, interaction quality, or pedagogical impact. The System Usability Scale (Brooke, 1996) is a general usability questionnaire that is applicable to chatbots. The BUS-15 instrument evaluates commercial chatbot usability; it has not been applied in educational contexts. To address this gap, mixed-methods research was conducted to evaluate the developed customised educational chatbot.

Participants

We recruited 37 upper-secondary school students aged 15–16 years (19 males, 18 females), randomly selected from one secondary school in Taiwan. The test was anonymous, and no personal data were collected. Participants volunteered during a scheduled IT lesson, using the school's internet and Chromebooks. The 90-minute session included three phases: a 20-minute introduction to the study goals, a 50-minute hands-on exploration of the chatbot's main functions via guided learning materials, and a 20-minute paper-based BUS 15 questionnaire. The study did not require a control group; thus, all students experienced the same condition, and informed consent was obtained beforehand. Demographic data, including age, sex, prior generative AI experience, and usage frequency, were recorded to contextualise the findings. With 37 participants, the sample size was sufficient to reveal key usability issues, considering that major problems typically surface within the first few dozen users (Faulkner, 2003).

Instrument and Procedures

The AI chatbot was developed using Botpress Cloud Studio (build 2024-11-19) on an Ubuntu 22.04 LTS desktop computer equipped with an Intel i7-7700 CPU and 32 GB RAM. Intent recognition was handled by the platform's continuously updated generative NLU model, eliminating the need for manual retraining. The system is automatically upgraded with the latest examples. Five primary flows (Login, Teacher's task, Parent's access, Student area, Questionnaire) were implemented, creating a total of 74 nodes. Each AI node ran the GPT-4o-Mini language model (August 2024 release), configured to produce responses that balanced creativity and consistency (a "temperature setting" of 0.6). The platform integrated the OpenAI-Botpress API (version 15.0.1), enabling cloud-based development. To evaluate the tool's usability and effectiveness, the BUS-15 questionnaire was employed. This standardised instrument offers a systematic framework for assessing user experience based on key dimensions: ease of use, efficiency, and satisfaction. Responses were measured using a Likert scale to capture nuanced perceptions of students' interactions with the chatbot. The questionnaire consists of 15 concise, age-appropriate items suitable for 15–16-year-old students, with a format brief enough to maintain engagement during a classroom activity.

A distinctive aspect of this study is the application of a commercial usability instrument within an educational context. The BUS-15 provides a structured methodology that ensures consistent measurement across different chatbot systems. Moreover, it addresses both functional and emotional aspects of user experience, offering a comprehensive view of student–AI interaction. Table 1 outlines the structure of the instrument.

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Table 1
BUS 15 Questionnaire

Factor	Item
1 - Perceived accessibility to chatbot functions	1. The chatbot function was easily detectable. 2. It was easy to find the chatbot.
2 - Perceived quality of chatbot functions	3. Communicating with the chatbot was clear. 4. I was immediately made aware of what information the chatbot could give me. 5. The interaction with the chatbot felt like an ongoing conversation. 6. The chatbot was able to keep track of context. 7. The chatbot was able to make references to a website or service when appropriate. 8. The chatbot could handle situations in which the line of conversation was not clear. 9. The chatbot's responses were easy to understand.
3 - Perceived quality of conversation and information provided	10. I found that the chatbot understood what I wanted and helped me to achieve my goal. 11. The chatbot gave me the appropriate amount of information. 12. The chatbot only gave me the information I needed. 13. I felt that the chatbot's responses were accurate.
4 - Perceived privacy and security	14. I think that the chatbot informed me of any possible privacy issues.
5 - Time response	15. The response time was short.

Source: Borsci et al. (2020)

Data Analysis

Findings from the BUS-15 questionnaire were systematically analysed to determine the internal consistency of the instrument in the context of an educational chatbot. I examined recurring themes in open-ended responses and statistically assessed Likert-scale ratings. Patterns in user feedback informed recommendations for system improvements and enhancements to the overall user experience. This dual-method approach enabled the identification of both strengths and weaknesses in the chatbot interactions. By integrating quantitative and qualitative data, I ensured a balanced evaluation that captured a diverse range of student perspectives. The BUS-15 remains a reliable instrument for assessing chatbot usability and guiding future developments. In this pilot study, the questionnaire functioned solely as a data-collection tool. No psychometric validation procedures (e.g., factor analysis or item-level diagnostics) were conducted, as scale development was not the primary objective. Descriptive statistical analysis was used to summarise and explore the distribution of student responses. This analysis facilitated a broad understanding of user sentiment and general trends, while helping to identify areas for potential improvement. Internal consistency metrics were calculated (Cronbach, 1951). Item-total correlation was excluded, as the aim was exploratory rather than diagnostic. Cronbach's alpha was deemed appropriate given that the BUS-15 is in English, while students' primary language is Traditional Chinese. This allowed an assessment of response consistency, cultural adaptation, and potential language barriers. Responses were processed using Python, selected for its simplicity, robust statistical libraries (e.g., Pandas, NumPy), and cross-platform compatibility. The anonymised dataset containing student responses is available via Zenodo (Fekets, 2025), thereby supporting transparency and reproducibility in future analyses. Finally, to explore deeper patterns and insights from the BUS-15 data, students were asked to respond to open-ended questions at the conclusion of the instrument. Table 2 presents the structure of this inquiry.

Table 2
Open-Ended Questions

Factor	Related question
Accessibility	How easy or difficult was it to start using the chatbot?
Function quality	How fun and interesting were the chatbot's tasks?
Conversation quality	How natural or robotic were the chats?
Privacy and security	Did you feel safe and comfortable using the chatbot?
Time response	How fast did the chatbot respond to you?

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

The open-ended responses were not analysed in depth, as the focus of the study was on quantitative usability trends derived from the Likert-scale items.

Research Results

To address the research questions, an AI chatbot was developed using the Agile methodology (Beck et al., 2001), tailored to support physics-focused STEM education. The system includes distinct modes for students, teachers, and parents, with access permissions and functionalities assigned according to predefined user privileges (see Figure 1). From Figure 2 to Figure 8, the developed menu and function list are presented.

Figure 1
Roles of Teachers, Parents, Students, and Admin

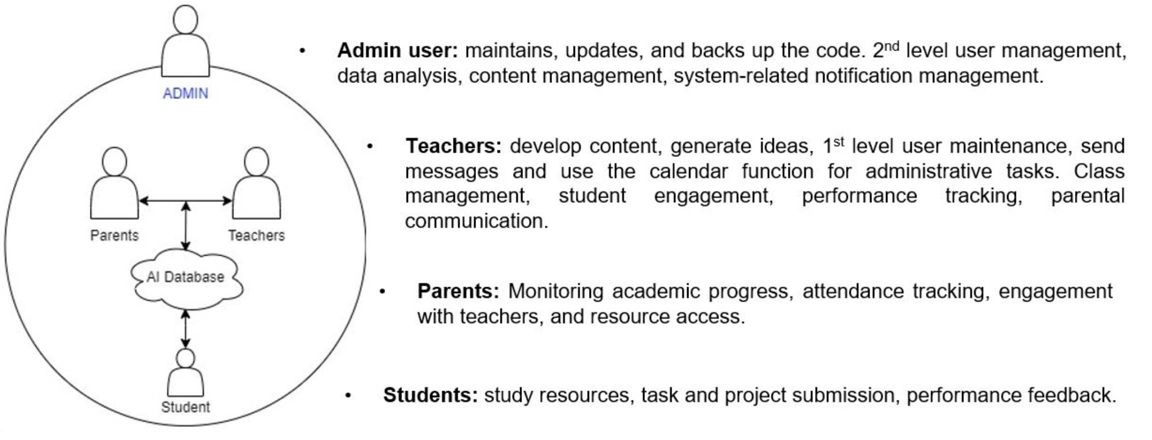
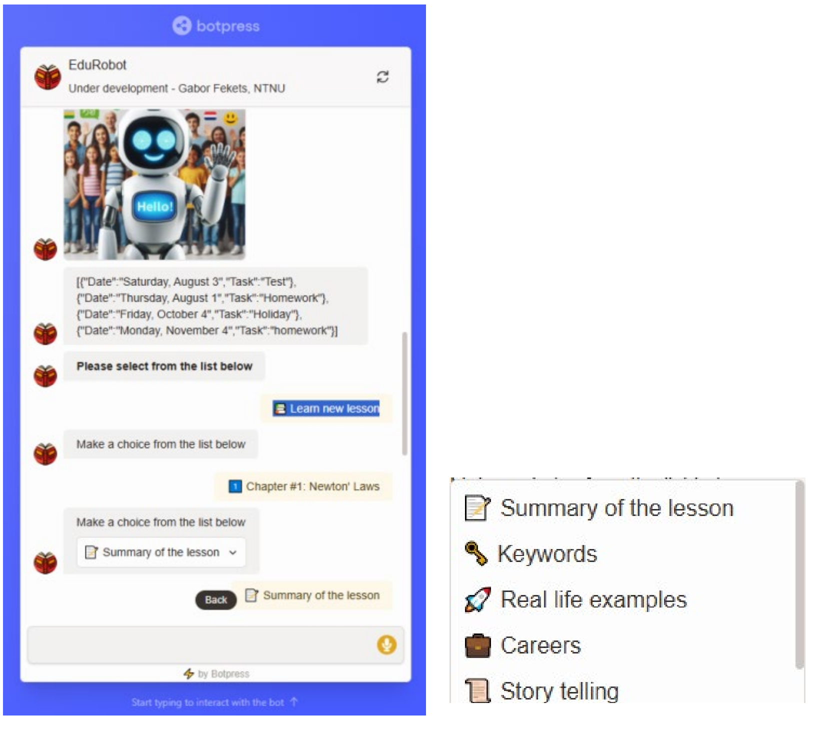


Figure 2
After Logging In, Students Can Select the Type of Activity Within the Task-Selection Interface.



https://doi.org/10.33225/jbse/25.24.878

Journal of Baltic Science Education, Vol. 24, No. 5, 2025

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Figure 3
The Lesson Summary Panel Displays a Brief Overview of the Session's Key Points.

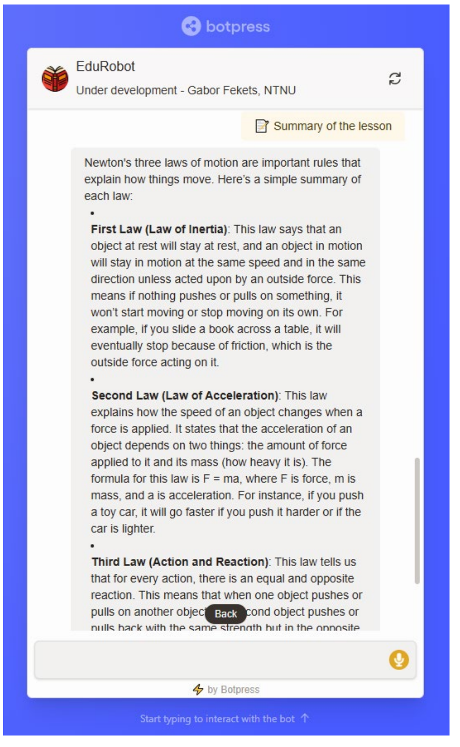
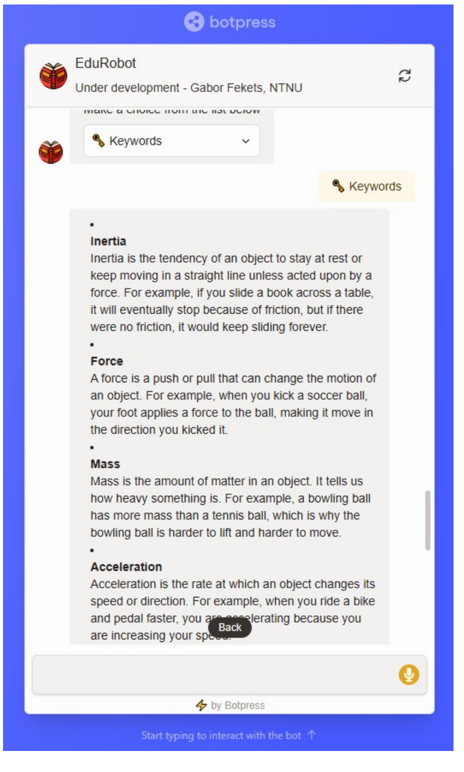


Figure 4
AI-Generated Content Produced by the Keywords Function, Illustrating Automated Keyword-Based Summarisation.



https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

The "Storytelling" session was generated by the AI system and presented a short narrative based on Newton's Law. After reading the story, students responded to questions designed to reinforce comprehension and retention of key concepts.

Figure 5
AI-Generated Storytelling Task Based on Newton's Law

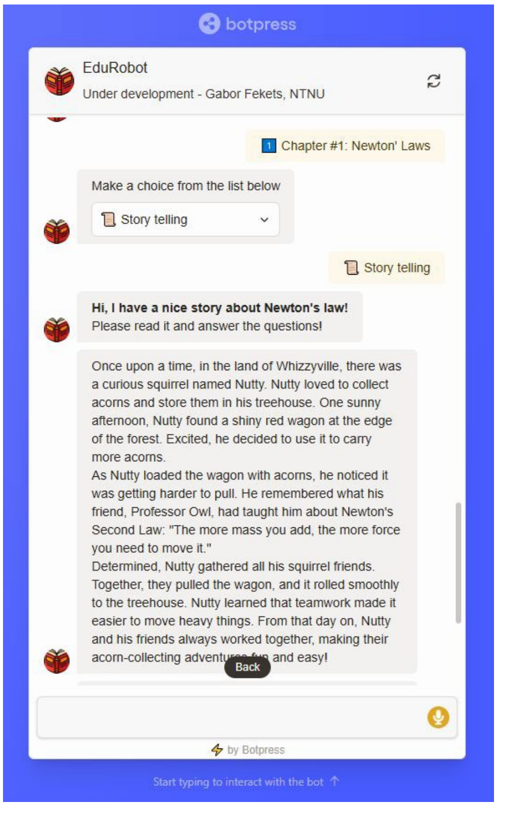
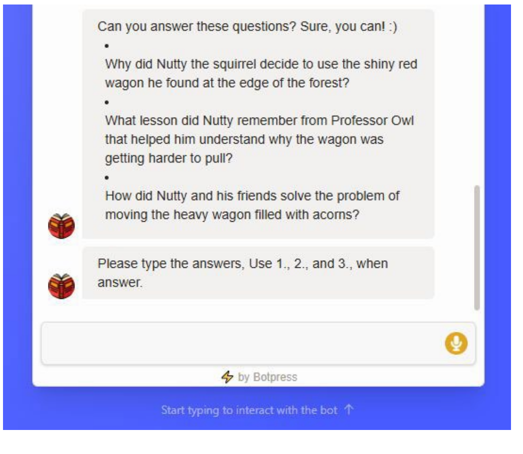


Figure 6 presents the comprehension questions based on the short story, while Figure 8 displays the scoring interface, providing immediate feedback on student responses. The underlying code also generated scores and tracked students' performance on the selected topic.

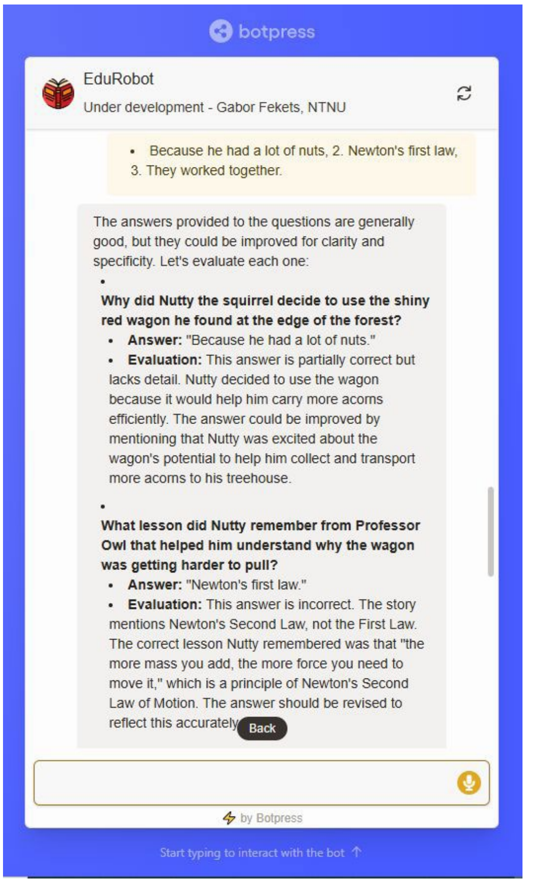
Figure 6
Storytelling Questions Generated by the AI System

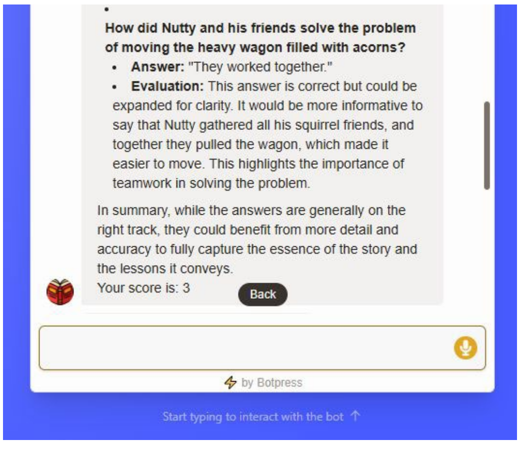


https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Figure 7
AI-Generated Scoring and Feedback Based on Student Responses

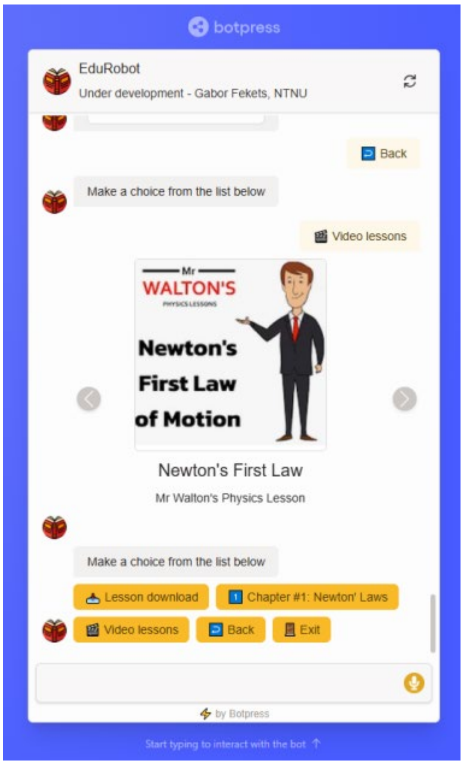




https://doi.org/10.33225/jbse/25.24.878

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Figure 8
Integration Feature Allowing Teachers to Embed YouTube Videos as Visual Content Within the Lesson Interface.



Qualitative and Quantitative Results

Descriptive statistics were computed for all questionnaire items to examine central tendency, variability, and distributional properties of the responses. The number of valid observations (*n*) varied slightly across items due to a small number of missing entries, ranging from 36 to 37 responses per item. Table 3 reports the calculated means, standard deviations, minimum and maximum scores, along with skewness and kurtosis coefficients for each item.

Table 3
Descriptive Statistics for Questionnaire Items

Item	*n*	*M*	*SD*	Min	Max	Skewness	Kurtosis
Q1	36	3.25	1.13	1	5	-0.02	-0.71
Q2	37	3.38	1.06	1	5	-0.52	-0.41
Q3	37	3.57	1.07	1	5	-0.32	-0.61
Q4	37	3.05	0.91	1	5	0.12	-0.32
Q5	37	2.73	1.1	1	5	-0.09	-0.9
Q6	37	2.35	1.27	1	5	0.62	-0.63
Q7	36	3.25	1.05	1	5	-0.51	-0.15
Q8	37	3.35	0.82	2	5	0.18	-0.45
Q9	37	2.62	1.06	1	5	-0.04	-0.69

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Item	*n*	*M*	*SD*	Min	Max	Skewness	Kurtosis
Q10	37	2.14	0.92	1	4	0.17	-1.02
Q11	37	3.16	1.01	1	5	-0.17	-0.41
Q12	37	2.51	1.1	1	5	0.16	-0.81
Q13	37	3.08	1.23	1	5	-0.16	-0.83
Q14	36	3.31	1.14	1	5	-0.15	-0.78
Q15	36	2.78	0.93	1	5	0.24	-0.45

Mean scores ranged from 2.73 (Q5) to 3.57 (Q3), reflecting moderate levels of agreement or endorsement depending on the content of each item. Standard deviations fell between 0.82 and 1.27, indicating acceptable variability in participants' responses. Skewness values ranged from –0.52 to +0.12, implying near-symmetric distributions with a slight leftward skew in some items. All kurtosis values were negative (–0.90 to –0.32), indicating platykurtic distributions with fewer extreme observations than expected under normality. Collectively, these metrics support the assumption of approximate normality, permitting the use of parametric analysis techniques. To evaluate internal reliability, Cronbach's alpha was calculated using complete responses only (i.e., excluding cases with missing data). The resulting coefficient was 0.709, representing acceptable internal consistency. This implies that the questionnaire items collectively measure a coherent construct, although some scope for refinement or item-level optimisation may remain. Overall, the descriptive results reveal balanced and varied student responses. The internal consistency of the instrument further supports its use for inferential analysis and scale-based interpretation. Table 4 summarises the average factor scores derived from the questionnaire.

Table 4
Average Scores for Each Factor

Factor	Number of items	*M*	*SD*	Evaluation
Accessibility	2	3.34	1.09	Strength
Function quality	7	2.99	1.05	Moderate performance
Conversation quality	4	2.72	1.04	Weak point
Privacy and security	1	3.31	1.14	Strength
Time response	1	2.78	0.93	Weak point

Across all questionnaire items, four missing responses were observed. Analyses were conducted using all available data for each item (pairwise inclusion), while reliability assessments were restricted to complete cases (listwise deletion, $n = 33$). Participants also responded to five open-ended questions aimed at exploring their subjective experiences with the autonomous conversational agent. These qualitative responses were not analysed using formal metrics but were employed as illustrative examples to contextualise the quantitative results. Students found accessing the chatbot (M = 3.34) via QR code intuitive, particularly on Chromebook devices. Typical comments included, "Using the QR code and entering the bot was easy," and "I also use QR codes on my phone frequently; it's easy." Function quality received a mean score of 2.99. Some students stated, "The bot's activities were useful, but I wish there were more of them," and "It is easier to learn with this bot." Conversation quality recorded the lowest average score (M = 2.72), with students frequently noting the challenge of navigating English-language content. Representative comments included, "My English is not so good; I read all of the sentences three or four times," and "I don't like typing too much, tapping is easier." Privacy and security (M = 3.31) did not raise concerns. Students completed the login process appropriately and did not report the collection of personal data. In one instance, a student attempted to access another account, but the system's use of unique credentials prevented unauthorised access. Response time (M = 2.78) was the second-lowest-rated factor. For students unfamiliar with chatbot interactions, delays in response attributed to AI processing requirements were occasionally noted. Some system messages also contributed to wait times, potentially stemming from backend limitations. In summary, these insights

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

contextualise the BUS-15 scores, reveal practical limitations, and offer guidance for refining chatbot functionality to enhance the user experience in future educational applications.

Discussion

With the rapid evolution of information technology, the adoption of AI tools in education is both inevitable and increasingly essential for enhancing traditional instructional practices. Among these tools, chatbots, an especially promising application of generative AI, have demonstrated considerable potential to foster student engagement in STEM disciplines. Mai et al. (2024) identified several limitations and threats associated with ChatGPT through a SWOT analysis. One notable limitation of general-purpose AI systems lies in their broad, non-specialised instructional outputs. The present study addressed this concern by designing a subject-focused chatbot tailored to physics education. Furthermore, the bot can be adapted across diverse subjects.

Previous research has highlighted additional limitations, including excessive response length (Chaudhry et al., 2023), content complexity (Stojanov, 2023), and minimal motivational impact on students facing challenging tasks (Yilmaz & Karaoglan Yilmaz, 2023). To counter these issues, the developed chatbot integrated a performance-based reward system, allowing students to earn points for their responses. This feature is designed to increase short-term engagement and promote task completion.

Concerns regarding diminished critical thinking, raised by Mohamed (2024) and Sallam et al. (2023), are particularly relevant in STEM education, where analytical reasoning forms a core component of curriculum goals. To foster such skills, the chatbot incorporates interactive questioning and contextualised real-life examples that encourage deeper cognitive engagement with physics concepts.

Despite its strengths, the chatbot cannot replicate the experiential learning and innovation essential to applied sciences education. While capable of supporting concept acquisition and foundational understanding, it does not substitute for hands-on experimentation or the nuanced communicative presence of a live teacher, including non-verbal cues and tone. As such, the chatbot should be considered a supplementary tool, a knowledge-rich, interactive support system, but not a standalone teaching solution.

An important achievement of this project was the use of the Botpress framework's Personality Agent, which delivers differentiated responses based on user age and prior knowledge. This feature enhances relevance and accessibility for diverse student populations. However, students with significantly lower proficiency levels may still face challenges in comprehension, underscoring the need for additional assessment-driven personalisation mechanisms. The performance-scoring system embedded within the chatbot provides useful motivational incentives, although its long-term effectiveness remains uncertain. Authentic motivation in STEM often arises from inquiry, discovery, and experiential learning, dimensions that cannot yet be fully replicated by autonomous systems. A potential drawback of AI-driven interaction is the tendency for students to seek immediate answers without pursuing deeper inquiry. This behaviour may limit their ability to examine multi-causal phenomena or develop holistic problem-solving skills. Encouraging critical evaluation and exploration from multiple angles is therefore vital.

In terms of strengths, the chatbot aligns well with contemporary educational goals. It offers personalised learning paths (Chan & Hu, 2023), delivers rapid responses (Limna et al., 2023), supports creative idea generation (Akiba & Fraboni, 2023; Liang et al., 2023), and enhances learning outcomes (Wandelt et al., 2023).

Future research opportunities include expanding the chatbot's feature set to support engagement and academic achievement further. Studies could evaluate whether AI-generated visuals enhance user experience, or whether performance-linked assessment contributes to improved learning outcomes, problem-solving capabilities, and knowledge retention. Integrating chatbot programming into IT and STEM curricula could also reinforce 21st-century skills. From a teacher's perspective, examining attitudes toward chatbot development and usage is equally valuable. Embedding pedagogical models such as Bloom's taxonomy (Ravichandran & Virgin B, 2024) may offer an effective framework for developing chatbot-based assessments.

To date, the system has not been introduced to educational stakeholders. Further review is necessary to evaluate its practical application and classroom readiness. Based on deployment outcomes, future iterations may require updates informed by feedback from students, teachers, and parents. Resource limitations, including free-tier access, operating-system compatibility, and emoji rendering, also restricted broader testing.

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

Qualitative Metrics

Next, the results were compared to those of the original BUS-15 questionnaire (Table 5).

Table 5

Comparison of Average Scores for Each Factor Between the Current Study and the Original Questionnaire

Factor	This study	Borsci et al. (2022)	Difference
Accessibility	3.34	3.62	-0.28
Function quality	2.99	3.78	-0.79
Conversation quality	2.72	3.6	-0.88
Privacy and security	3.31	3.52	-0.21
Time response	2.78	3.3	-0.52

Compared to the original findings of Borsci et al. (2022), I found lower average scores across all five BUS-15 factors, with differences ranging from −0.21 to −0.88. Several contextual variables may account for this disparity. First, participants interacted with the chatbot in English, which was not their native language. Their age and limited proficiency may have adversely affected the perceived conversation quality. Second, cultural response patterns likely played a role: respondents from Southeast Asia are generally inclined to avoid extreme Likert-scale ratings (Harzing, 2006). Third, the study was conducted in a live classroom setting, which may have elicited more critical feedback compared to evaluations in more controlled environments. Finally, variability in prior exposure to AI tools, along with the relative developmental maturity of the chatbot, could have influenced participants' usability assessments.

General Discussion

This pilot study provides valuable information on the usability of the proposed AI system and its initial phase of development. However, future research should repeat the analysis with a larger and more diverse participant base to enhance generalisability. A key operational metric is the number of interactions between students and the system: although classroom planning anticipated 1,200 to 1,500 chatbot queries, usage logs showed approximately 3,500 question–answer pairs, reflecting unexpectedly high engagement. Future studies should incorporate a dedicated interaction counter linked to individual users, rather than relying solely on session time, which is prone to inaccuracy when users neglect to log out.

This research had several limitations. As a pilot with a modest sample size, it may not have captured the full scope of usability challenges or design opportunities, reinforcing the need for larger-scale replication. The BUS-15 questionnaire was administered within a single 90-minute session, and the scope was limited to student-facing content. Although the chatbot also includes roles for parents and teachers, such as academic monitoring, calendar management, and parent–teacher interaction, these functions were not examined. Language barriers further complicated interaction quality. Operating solely in English, the chatbot presented comprehension challenges for students unfamiliar with the language, resulting in misinterpretations and repeated queries. Enhancing conversation quality requires systematic validation of translations and revision of instructional phrasing. In addition, the user guide should be updated to provide clearer, step-by-step instructions and contextual examples of system functionality.

Conclusions and Implications

This study underscores the potential of generative AI, particularly customised chatbots, as a key component in enhancing student–computer interaction within STEM education. The increasing demand for personalised and interactive learning experiences highlights the importance of developing accessible, student-friendly applications. The chatbot, developed using the Botpress platform, benefited from a flexible and robust framework suitable for educational deployment. Such tools can effectively support student engagement and complement existing teaching practices.

However, the analysis revealed that chatbot development is not universally accessible to educators. A lack of programming expertise and technical background may limit participation. Therefore, collaboration between

https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

educators and IT specialists is essential during development, as is continued training and support to build technological confidence.

The study highlights new curricular opportunities in programming education. First, further empirical research is needed to explore the use of AI-powered chatbots in teaching programming languages. While current tools can generate syntactically accurate code across most languages, they fall short in fostering computational thinking and debugging skills. A well-designed, customised chatbot could help identify student knowledge gaps and cultivate deeper analytical reasoning. Second, the integration of chatbot development into IT curricula represents a valuable addition to traditional programming instruction. Although foundational languages such as Python, C++, and JavaScript remain critical, competency in chatbot design offers a unique and increasingly relevant skillset for the workforce.

AI-based applications also show promise for boosting student interest in STEM subjects. This small-scale pilot serves as an initial step toward embedding AI tools within pedagogical frameworks. The goal is not to deliver cutting-edge systems for their own sake but to support learner growth, motivation, and achievement. Realising this goal requires expanded research, functional enhancements, and sustained pedagogical integration to prepare students adequately for the digital future of education.

Acknowledgements

The author would like to express a sincere gratitude to the supervisor at National Taiwan Normal University for valuable guidance and continuous support throughout the research project. The author also thanks the participating students and the school administration for their collaboration in the pilot study. Grammarly software was used to support the preparation of this manuscript and to enhance the quality of the English language. To promote transparency and reproducibility, the anonymised dataset used in this study is openly available via Zenodo: https://doi.org/10.5281/zenodo.15767971

References

Akiba, D., & Fraboni, M. C. (2023). AI-supported academic advising: Exploring ChatGPT's current state and future potential toward student empowerment. *Education Sciences*, *13*(9), Article 885. https://doi.org/10.3390/educsci13090885

Bazelais, P., Lemay, D. J., & Doleck, T. (2024). User acceptance and adoption dynamics of ChatGPT in educational settings. Eurasia Journal of Mathematics, Science and Technology Education, 20(2), Article em2393. https://doi.org/10.29333/ejmste/14151

Beck, K., Beedle, M., Bennekum, A. van, Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R. C., Mellor, S., Schwaber, K., Sutherland, J., & Thomas, D. (2001, February). Manifesto for Agile Software Development. Retrieved July 4, 2024, from https://agilemanifesto.org/

Bettayeb, A. M., Abu Talib, M., Sobhe Altayasinah, A. Z., & Dakalbab, F. (2024). Exploring the impact of ChatGPT: Conversational AI in education. *Frontiers in Education*, *9*, Article 1379796. https://doi.org/10.3389/feduc.2024.1379796

Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The chatbot usability scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing*, *26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability Evaluation in Industry* (pp. 207–212). CRC Press. https://doi.org/10.1201/9781498710411-35

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 43. https://doi.org/10.1186/s41239-023-00411-8

Chaudhry, I. S., Sarwary, S. A. M., El Refae, G. A., & Chabchoub, H. (2023). Time to revisit existing students' performance evaluation approach in higher education sector in a new era of ChatGPT — a case study. *Cogent Education*, *10*(1), Article 2210461. https://doi.org/10.1080/2331186X.2023.2210461

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

El Azhari, K., Hilal, I., Daoudi, N., & Ajhoun, R. (2023). SMART chatbots in the E-learning domain: A systematic literature review. *International Journal of Interactive Mobile Technologies (iJIM)*, *17*(15), 4–37. https://doi.org/10.3991/ijim.v17i15.40315

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, *35*(3), 379–383. https://doi.org/10.3758/BF03195514

Federici, S., De Filippis, M. L., Mele, M. L., Borsci, S., Bracalenti, M., Gaudino, G., Cocco, A., Amendola, M., & Simonetti, E. (2020). Inside Pandora's box: A systematic review of the assessment of the perceived quality of chatbots for people with disabilities or special needs. *Disability and Rehabilitation: Assistive Technology*, *15*(7), 832–837. https://doi.org/10.1080/17483107.2020.1775313

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Fekets, G. (2025). *Anonymised BUS-15 dataset from a STEM chatbot pilot study (N = 37, Taiwan, 2025)* (Version 1.00) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.15767971

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323–327. https://doi.org/10.1016/j.intcom.2010.04.004

Ghayoomi, M. (2023). Applications of chatbots in education: In M. A. Kuhail, B. Abu Shawar, & R. Hammad (Eds.), *Advances in Web Technologies and Engineering* (pp. 80–118). IGI Global. https://doi.org/10.4018/978-1-6684-6234-8.ch004

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, *6*(2), 243–266. https://doi.org/10.1177/1470595806066332

Hasan, N., Polin, J. A., Ahmmed, Md. R., Sakib, Md. M., Jahin, Md. F., & Rahman, Md. M. (2024). A novel approach to analyzing the impact of AI, ChatGPT, and chatbot on education using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *13*(4), 2951–2958. https://doi.org/10.11591/eei.v13i4.7158

Liang, Y., Zou, D., Xie, H., & Wang, F. L. (2023). Exploring the potential of using ChatGPT in physics education. *Smart Learning Environments*, *10*(1), Article 52. https://doi.org/10.1186/s40561-023-00273-7

Limna, P., Kraiwanit, T., Jangjarat, K., Klayklung, P., & Chocksathaporn, P. (2023). The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning & Teaching*, *6*(1), 64–74. https://doi.org/10.37074/jalt.2023.6.1.32

Luzano, J. F. P. (2024). Reshaping mathematics instruction via impact of AI chatbots on secondary education pre-service teachers. *International Journal of Studies in Education and Science*, *5*(3), 233–245. https://doi.org/10.46328/ijses.97

Mai, D. T. T., Da, C. V., & Hanh, N. V. (2024). The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. *Frontiers in Education*, *9*, Article 1328769. https://doi.org/10.3389/feduc.2024.1328769

Mishra, R., & Varshney, D. (2024). Comprehensive analysis of human and AI task allocation in the education sector: Defining futuristic roles and responsibilities. *World Journal of Advanced Research and Reviews*, *22*(3), 1883–1893. https://doi.org/10.30574/wjarr.2024.22.3.1949

Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, *29*(3), 3195–3217. https://doi.org/10.1007/s10639-023-11917-z

Neji, W., Boughattas, N., & Ziadi, F. (2023). Exploring new AI-based technologies to enhance students' motivation. *Issues in Informing Science and Information Technology*, *20*, 95–110. https://doi.org/10.28945/5149

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, *55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, Article 100033. https://doi.org/10.1016/j.caeai.2021.100033

Opesemowo, O. A. G., & Adewuyi, H. O. (2024). A systematic review of artificial intelligence in mathematics education: The emergence of 4IR. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(7), Article em2478. https://doi.org/10.29333/ejmste/14762

Ravichandran, K., & Virgin B, A. (2024). Bloom's taxonomy categories in the economy of literature teaching-learning process. *International Research Journal of Multidisciplinary Scope*, *05*(03), 721–727. https://doi.org/10.47857/irjms.2024.v05i03.0827

Sallam, M., Salim, N., Barakat, M., & Al-Tammemi, A. (2023). ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J*, *3*(1), Article e103. https://doi.org/10.52225/narra.v3i1.103

Saxena, A., & Doleck, T. (2023). A structural model of student continuance intentions in ChatGPT adoption. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(12), Article em2366. https://doi.org/10.29333/ejmste/13839

Schei, O. M., Møgelvang, A., & Ludvigsen, K. (2024). Perceptions and use of AI chatbots among students in higher education: A scoping review of empirical studies. *Education Sciences*, *14*(8), Article 922. https://doi.org/10.3390/educsci14080922

Segovia-García, N. (2024). Optimizing student support. A review of the use of AI chatbots in higher education. *European Public & Social Innovation Review*, *9*, 1–20. https://doi.org/10.31637/epsir-2024-324

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 35. https://doi.org/10.1186/s41239-023-00404-7

Valério, F. A. M., Guimarães, T. G., Prates, R. O., & Candello, H. (2018). Chatbots explain themselves: Designers' strategies for conveying chatbot features to users. *Journal on Interactive Systems*, *9*(3), 61–79. https://doi.org/10.5753/jis.2018.710

Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1997). PARADISE: A framework for evaluating spoken dialogue agents. *Proceedings of the 35th Annual Meeting on Association for Computational Linguistics*, 271–280. https://doi.org/10.3115/976909.979652

Wandelt, S., Sun, X., & Zhang, A. (2023). AI-driven assistants for education and research? A case study on ChatGPT for air transport management. *Journal of Air Transport Management*, *113*, Article 102483. https://doi.org/10.1016/j.jairtraman.2023.102483

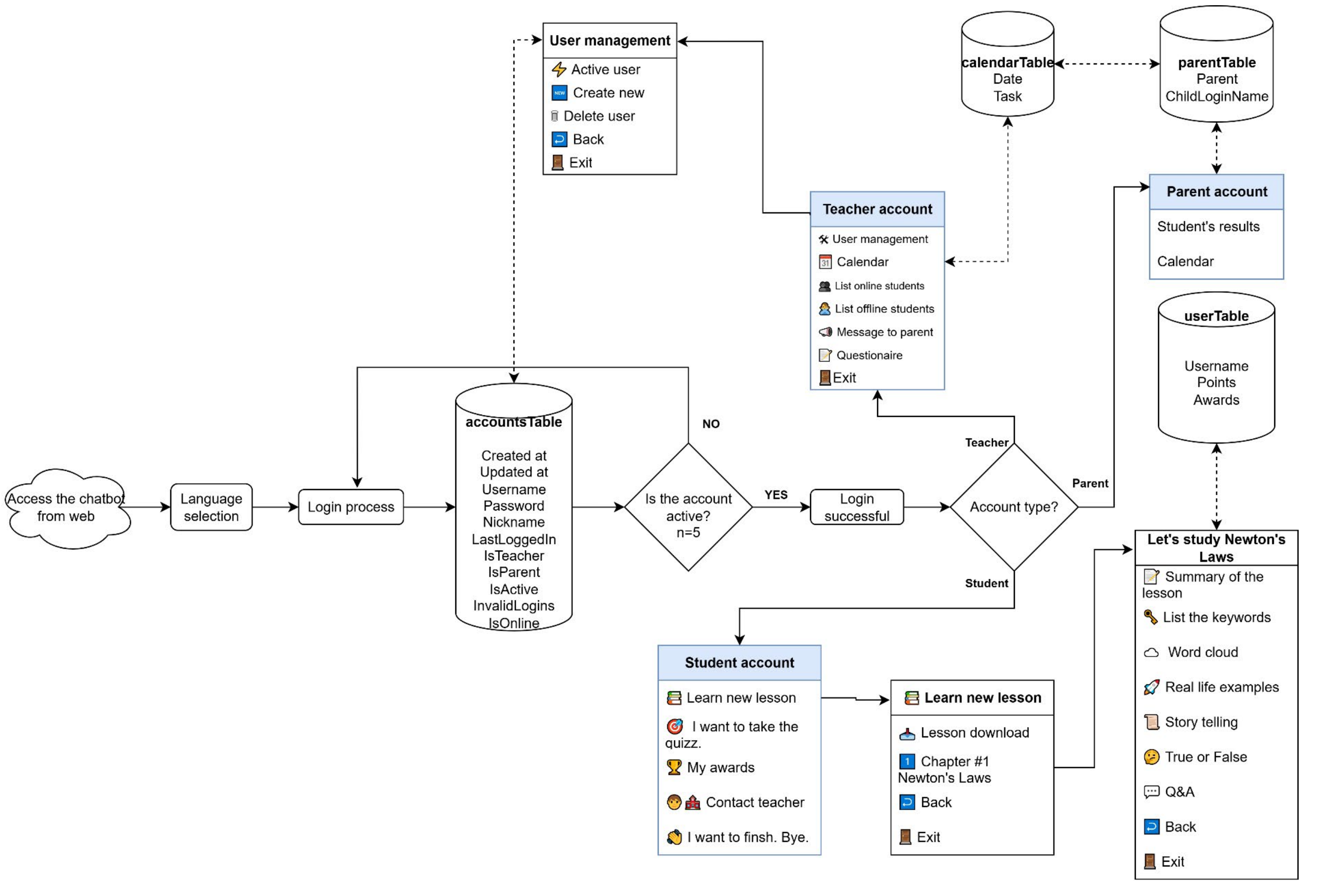
Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), Article em2286. https://doi.org/10.29333/ejmste/13272

Yilmaz, R., & Karaoglan Yilmaz, F. G. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy, and motivation. *Computers and Education: Artificial Intelligence*, *4*, Article 100147. https://doi.org/10.1016/j.caeai.2023.100147

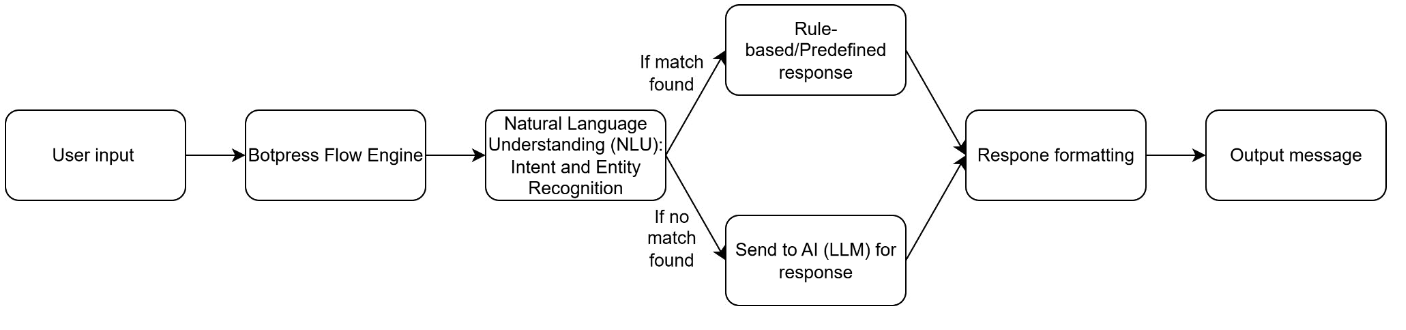
ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Appendix

The Developed Chatbot's Functions and Services



The Botpress GUI and the AI Model Integration



https://doi.org/10.33225/jbse/25.24.878

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Received: *April 02, 2025* Revised: *June 29, 2025* Accepted: *August 15, 2025*

Cite as: Fekets, G. (2025). Development of an artificial intelligence-supported Chatbot as an interactive learning platform in STEM education: Exploring usability and student experience. *Journal of Baltic Science Education*, *24*(5), 878–893. https://doi.org/10.33225/jbse/25.24.878

Gábor Fekets PhD Student of International Doctoral Program in Integrative STEM Education, College of Technology and Engineering, National Taiwan Normal University, Taiwan.
E-mail: 81178001h@ntnu.edu.tw
ORCID: https://orcid.org/0009-0007-5584-6520